

Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) EP 0 838 751 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:  
29.04.1998 Bulletin 1998/18

(51) Int Cl.<sup>6</sup>: G06F 3/033

(21) Application number: 97305193.1

(22) Date of filing: 14.07.1997

(84) Designated Contracting States:  
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC  
NL PT SE  
Designated Extension States:  
AL LT LV RO SI

- Flynn, Michael John  
Cambridge CB2 2RN (GB)
- Carter, Kathleen A.  
Cambridge CB5 8HH (GB)
- Newman, William  
Cambridge (GB)

(30) Priority: 12.07.1996 GB 9614837

(71) Applicant: RANK XEROX LIMITED  
Marlow, Buckinghamshire SL7 1YL (GB)

(74) Representative: Reynolds, Julian David  
Rank Xerox Ltd  
Patent Department  
Parkway  
Marlow Buckinghamshire SL7 1YL (GB)

(72) Inventors:  
• Wellner, Pierre  
Allenhurst, NJ07711 (US)

(54) Interactive desktop system with adjustable image capture

(57) An interactive desktop system comprising a work surface, a video projector for displaying images in a display area, an camera for capturing images present on the work surface, and an image processing computer coupled to the projector and to the camera, and adjustment means for adjusting, in response to user action, the field of view of the camera. In the preferred embodiment, the adjustment means comprises a frame including a housing in which the camera is disposed, the frame being movable (in a manner similar to that of a shade of

a desk lamp) whereby the field of view of the camera is altered. In another embodiment, the image processing computer is adapted for modifying, in response to a user input, the dimensions of the display area and/or the resolution of the images displayed therein. The image processing computer may include means, responsive to a user input designating a position on the work surface, for causing movement of the housing whereby the said designated position is moved to within the field of view of the camera.

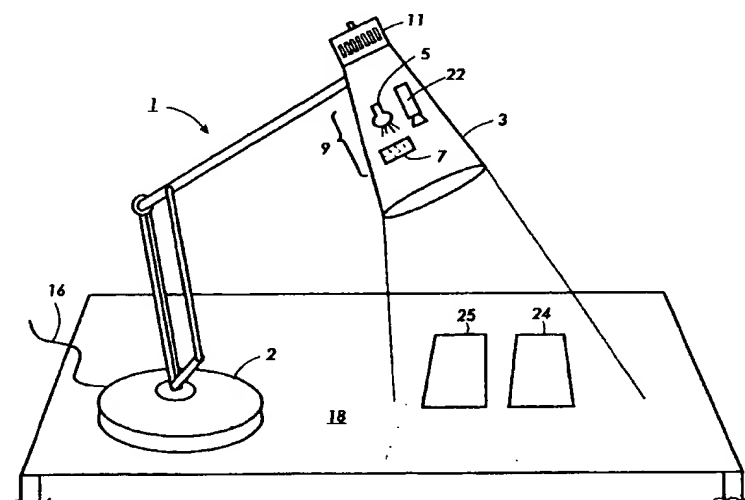


FIG. 1

## Description

This invention relates to a document processing system in which a user manipulates documents interactively via using a camera positioned above a desk, and more particularly to such a system having adjustable image capture.

It is known from EP-A-495 622 to use a camera-projector arrangement positioned above a desk, in order to enable a user to select functions to be performed by selecting items located within the field of view of the camera. A video camera or scanner is used to capture images of documents on a desk, and feedback information is displayed by means of a projection display. The functions include calculating and translating operations carried out on data (e.g., in a paper document) located on the desk.

EP-A-622,722 discloses a system for generating new documents from originals containing text and/or images employing e.g. a camera-projector system focused on a work surface, in conjunction with a copier or printer. In use, the camera captures various manual operations carried out by the user, e.g. by pointing with fingers and tapping on the surface on the text or images in an original paper document on the surface and representing manipulations of the text or images. Feedback to the user is provided by projection of an image onto the surface or onto the original, or using some other visual display.

A problem with the system of EP-A-622,722 is that its use of a camera and projector which are fixedly mounted on office furniture, bookshelves or the ceiling is that it is very bulky, difficult to set up, and difficult to move. For practical purposes, many potential users would find it impossible to set up over their existing desks. Another problem is the limited field of view of the camera and the fact that it is inconvenient to zoom it and point it at different locations on the desk.

A further problem with the system of EP-A-622,722 is that the projector has a limited resolution and a limited range, and there is a trade-off between the two. Some applications (e.g. fine sketching or handwriting) require high resolution per square inch, while other applications (e.g. Architectural visualisation) require a large surface area. Any given projection must sacrifice one for the other. One aspect of this invention concerns an interaction technique and input device that addresses the area/resolution trade-off.

A further problem arises through the low resolution of the images produced by the video cameras of the above-described systems (which are based on current standard video technology) compared with that of printers and scanners. This leads to low accuracy when recognising the text of selections, and poor quality when printing selections. When the user is working with relatively small selections at a time (words or drawings less than a few inches across), then one approach to solving this problem is to zoom the camera to focus only on the

area of interest. The problem then is how to control the camera. One approach is to have the user manually adjust the camera zoom before making a selection, but this requires an extra step, extra controls, and a monitor of some sort so that the user can see "through the camera". It would be desirable to have a system that could automatically control the camera to zoom in as tightly as possible on the area selected by the user.

The present invention provides interactive desktop system comprising a work surface, a display device for displaying images in a display area, an image capture device for capturing images present on the work surface, and processor means, coupled to the display device and said capturing means, and adjustment means for adjusting, in response to user action, the field of view of the image capture device.

The present invention further provides an interactive desktop system comprising a work surface, a display device for displaying images, an image capture device for capturing images present on the work surface, and processor means, coupled to the display device and said capturing means, a frame including a housing in which the image capture device is disposed, the frame being movable whereby the field of view of the image capture device is altered.

An advantage of this system is its compatibility with existing office equipment, and provision of a very natural interaction technique for controlling the zoom factor and where the camera and projector point (discussed further below in section B).

The present invention further provides an interactive desktop system comprising a work surface, a display device for displaying images in a display area, an image capture device for capturing images present on the work surface, and processor means, coupled to the display device and said capturing means, the processor means including means for modifying, in response to a user input, the dimensions of the display area and the resolution of the images displayed therein.

With existing technology, if the zooming and position of the projector are controllable at all by the user, the controls are independent from the size and view of the image projected. This forces the user to select a trade-off between size and resolution which once made, is difficult to change. The technique according to the invention makes it easy to move smoothly between all possible values of this trade-off, so the limitations of any particular position are less of a problem. This technique could be compared to the techniques of navigating in a virtual world (e.g. flight simulation) but is the opposite in a way. Instead of showing a different view, according to the user's position in the computer's virtual world, what is shown is a different view according to the computer's position in the user's real world.

The present invention further provides an interactive desktop system comprising a work surface, a display device for displaying images, an image capture device for capturing images present on the work surface,

and processor means, coupled to the display device and said capturing means, a frame including a housing in which the image capture device is disposed, the frame being movable, the processing means including means, responsive to a user input designating a position on the work surface, for causing movement of said housing whereby the said designated position is moved to within the field of view of said image capture device.

Preferably, the processing means includes: means for determining from a captured image whether the predetermined form is horizontally centred within the field of view of said image capture device, and if not, for causing the image capture device to pan until the predetermined form is horizontally centred; and/or means for determining from a captured image whether the predetermined form is vertically centred within the field of view of said image capture device, and if not, for causing the image capture device to tilt until the predetermined form is vertically centred.

Preferably, the processing means includes: means for determining from a captured image whether the predetermined form is fully visible within the field of view of said image capture device, and if so, for causing the image capture device to electronically zoom in until the predetermined form is no longer fully visible; and means for determining, after the predetermined form is no longer fully visible, from a captured image whether the predetermined form is fully visible within the field of view of said image capture device, and if not, for causing the image capture device to electronically zoom out until the predetermined form is just fully visible.

The predetermined form may comprise a selection rectangle shaped by a user, and said elements comprise the side of said rectangle.

This technique has the advantage that a user can simply select an area on the work surface and the image capture device (camera) is automatically panned, zoomed, and/or tilted into position so that the selected area is just within its field of view — enhancing resolution of the captured image in the selected area.

The display device preferably comprises a video projector. Alternatively, the display device may comprise a desk-mounted VDU, such as a CRT monitor, or a flat panel display, such as a LCD display. Preferably, the image capture device comprises a video camera.

Embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 shows a view of the overall system according to one embodiment of the invention;  
Figure 2 is a schematic block diagram of the computer used in the system of Fig. 1;  
Figure 3 illustrates an alternative embodiment employing mirror-based projection in accordance with the invention; and  
Figure 4 shows a further embodiment of the invention employing a lamp housing movable in three di-

mensions, for projecting a display of variable size and resolution;

Figure 5 shows a further embodiment, similar to that of Fig. 4, but maintaining the projected display in a different orientation;

Figure 6 is a flow chart illustrating the processing steps used in the systems of Figs 4 and 5;

Figure 7 is a view of the work surface illustrating a further embodiment of the invention accomplishing automatic pan/til/zoom of the camera to user-selected areas on the work surface; and

Figure 8 is a flow chart illustrating the processing steps used in the system of Fig. 7.

## 15 A. Basic System Configuration

The general system configuration according to the invention, employing a video camera and projector positioned above and focused on a work surface 18, is illustrated in Fig. 1. The system 1 takes the general form of a conventional desk lamp, with the camera 22 and video projector 9 (comprising bulb 5 and LCD matrix 7) being mounted within lamp housing 3. The base housing 2, which is suitably larger than that for a conventional lamp, provides a housing for computer hardware, for example having a standard microcomputer (PC) architecture. (Alternatively, the processing hardware may be packaged in a standard desktop PC housing, with the base housing 2 being similar to a standard desk lamp, and the camera 22 and projector 9 being coupled to appropriate ports of the desktop PC by suitable cables (not shown).) It will be appreciated by persons skilled in the art that the invention may be implemented using a computer running Windows™ and equipped with a frame grabber board and appropriate interfacing circuitry (see, for example, Jähne B., *Digital Image Processing*, Appendix B, Springer-Verlag, Heidelberg, 1991).

Alternatively, the hardware configuration described in detail in published European application EP-A-622,722, may be employed except that only interfacing circuitry is provided in housing 2, which is in turn connected to the remote processing system. The camera 22 and video projector 9 are then connected to image processing hardware described in detail in that application.

In another alternative, the housing may contain only the camera 22, with the system being implemented as described in commonly-owned GB patent application No. 9614694.9 (Agent's ref: R/96007), filed concurrently herewith.

In implementing the present invention, image processing techniques described in EP-A-622,722 may be used, modified, as appropriate and/or as described below.

The computer is coupled to a network cable 16, which in conjunction with a conventional internal driver card (not shown) and appropriate software (such as Netware, available from Novell Corp.) enable the computer

to form part of a local (LAN) or wide area network (WAN), as is well known in the art.

The lamp housing 3 is suitably arranged to be movable in three dimensions, exactly in the manner of a conventional angle-poise lamp, enabling the camera 22 and projector 9 to be moved towards and away from the desk surface 18 as desired for particular operations (in Fig. 1 the angle that the line of sight of the camera/projector makes with the vertical is exaggerated, for the purpose of illustration). As shown, a real paper document 24 is within the field of view of the camera 22, and an electronic document 25 is projected by video projector 9 on to the surface 18. At the top of the lamp housing 3 is further provided a rotary zoom control 11, similar to that provided on conventional SLR and video cameras, but electronically linked to the computer processor so as to enable the user to perform control of the zoom setting of the camera 22.

The hardware of the computer system 1 is schematically shown in Fig. 2. The configuration is well known to persons skilled in the art and, for example, is described in detail in *The Art of Electronics*, 2nd Edn, Ch. 10, P. Horowitz and W. Hill, Cambridge University Press, 1989. Stated briefly, the system comprises, connected to common bus 30, a central processing unit 32, memory devices including random access memory (RAM) 34, read only memory (ROM) 36 and disk, tape or CD-ROM drives 38, keyboard 12 (not shown), mouse 14 (not shown), printing, plotting or scanning devices 40, and A/D, D/A devices 42 and digital input/output devices 44 providing interfacing to external devices 46 such as camera 22, video projector 9 and the rest of the LAN (not shown).

In use, the video camera 22 captures images of the desk-surface 18 and these images are displayed in real-time by the video projector 9 on the desk surface 18.

The system must maintain a mapping between the pixels in the camera's field of view and the pixels of the display. This mapping will change as the angle of the lamp housing 3 with respect to the desk changes, so either this angle must be kept fixed, or a quick self-calibration procedure (see, for example, EP-A-622,722, or GB patent application 95 210 72.0) can be used.

A further embodiment of the invention is illustrated in Fig. 3(a). This embodiment is the same as the first-mentioned embodiment, except as described below. The system addresses a problem encountered with a system employing a camera 22 to scan documents lying on the desk 52, and a video projector 9 to display feedback information onto the desk surface 18, i.e. the integration of the video projector 9, which is necessarily bulky due to the size of its LCD array, projection light source and cooling fans, with the camera 22 that scans the desk.

The solution is to attach a mirror 50 to the underside of the camera housing, and to mount the projector 9 separately, pointing upwards at the mirror 50. Here, the projector 9 may take the form of a standard overhead pro-

jector having a lens 54 for use in adjusting focus; and the LCD panel 7 is disposed on the upper surface of the projector. The angle of inclination of the projector is exaggerated. As a result the image (having width P) is projected back down onto the same region (having width F) of the desk as is scanned by the camera. The projector may be positioned behind the desk, or under the desk projecting up through a hole, as shown in Fig. 3(a). The mirror 50 is hinged at A and a screw adjustment at B allows the angle of the mirror 50 to be adjusted so that the image is projected on the appropriate part of the desk surface 18.

The image must be reversed to counter the effects of reflection by a single mirror 50. Most projectors and LCD panels having this reversal capability built in. Also, steps may need to be taken to minimise the keystone (foreshortening) distortion and focusing problems caused by the image being projected at an angle to the vertical. Again, some self-contained projectors include the means to displace the lens with respect to the LCD array 7 that generates the image. This may be sufficient to eradicate keystone altogether, as shown in Fig. 3 (b), and at the same time to bring the entire image into focus. A small amount of keystone can be tolerated by the user without difficulty.

#### B. Projecting a display of arbitrary size, position and resolution

Figure 4 illustrates a further embodiment of the invention: this is exactly the same as the first-mentioned embodiment, except as described below.

In this embodiment, the lamp housing 3 is preferably restricted to being directed straight down (i.e. the angle of the optical axis of the camera 22 and projector 9 to the vertical is zero: it is perpendicular to the desk surface 18). (In an embodiment where this angle is variable (see, e.g. Fig. 1) an inverse projective transformation is required to produce an undistorted image).

In this embodiment, the lamp base 2 is configured such that the system (and therefore the lamp housing 3) can rotate about an axis 60.

Provision is made for sensing not only the vertical distance z from the lamp housing 3 to the desk surface 18, but the horizontal position of the lamp housing 3 in two dimensions. For this purpose, shaft encoders (which output electrical signals indicative of (change of) angular position) are provided at points C, D and E. It will be appreciated that alternative means may be used for measuring the distance z: an ultrasonic transducer such as are employed in autofocus cameras may be used, disposed on the lamp housing 3.

As seen in Fig. 4(b), which shows a plan view, using the swivel base 2 means that the display region 62 rotates as the system is swivelled. This is compensated for by applying the reverse rotation ( $-\theta$ ) to the image before displaying it.

(A further alternative embodiment is illustrated in

Fig. 5: this is the same as in Fig. 4, except that a more complex double linkage is used which maintains the display region 62 orthogonal to the desk).

In the embodiments of Figs. 4 and 5, as the distance  $z$  from the desk surface 18 varies, the focus of the projector 9 must be maintained. In accordance with this aspect of the invention, adjustment is made automatically based on measured changes in  $z$ .

Figure 6 is a flow chart of the processing steps involved in adjusting the displayed (projected) image following an alteration (change in  $x$ ,  $y$  or  $z$  in the co-ordinate system of Figs. 4 and 5) of the position of the lamp housing 3 by the user. In Fig. 6, allowance is made for varying  $\theta$  (Fig. 4), but for the embodiment of Fig. 5,  $\theta = 0$ .

As can be seen in Fig. 6, initially (step s2), the angles at the shaft encoders are sensed. If a change in the angle(s) is detected (step s4), computations are performed (based on well known geometric relations) to determine from the shaft encoder angles the projector's position ( $x$ ,  $y$ ,  $z$ , (and  $\theta$ )) in the 3-D space. If it is determined (step s8) that a change in vertical position ( $z$ ) has taken place, further calculations are performed, based on the new value of  $z$  and the relevant geometric relations, to compute (step s10) the size of the new display region. The focus for the projector is then adjusted (step s12) by a corresponding amount to ensure a sharp image. Next, at step s14, the position of the display region in the virtual display is computed, and then a rotation of  $-\theta$  is then performed to counteract any detected rotation  $\theta$ . Then (step s18), operations are performed to render the display region onto the full region of the display, and the rendered image displayed (step s20).

The system maintains a virtual projected surface that is of arbitrarily large size and arbitrarily high resolution. At any given moment the projector is displaying a portion of this surface at a particular resolution. The user is given the impression of arbitrary size and resolution because he or she can move the projector (or zoom it) closer to the desk and virtual objects stay the same size but are displayed with more pixels per inch. The size of the area can be made bigger by zooming (or moving) away, or the projector can be moved to the area that needs to be seen. This system is ideally suited to the system of Figs. 1, 4 and 5, where the user can "light up" any part of the virtual surface area desired by moving the lamp around. Then any area can be inspected to arbitrarily high resolution by moving the lamp closer. It is believed that this interaction technique would be familiar and very easy to learn. The only hardware required in addition is the sensors so the system knows how far it and its angle is from the surface it is projecting on.

#### C. Automatic pan/zoom to selections on the desktop.

A further embodiment of the invention will now be discussed with reference to Fig. 7: this is exactly the same as the previous embodiments, except as de-

scribed below. According to this embodiment, the camera 22 is provided with an electronically controllable zoom, and the system provides automatic control of the camera's zoom by projecting a feedback rectangle 66 that is also usable (detectable) by the system itself. The projected selection rectangle 66 is suitably of a predetermined grey level enabling its position to be determined. Alternatively, within the individual sides of the selection rectangle may be included coded information (e.g., glyph codes: see, for example, US-A-5,168,147; US-A-5,091,066; and US-A-5,128,525).

According to this technique, when the system can "see" the (whole of) the selection rectangle 66 that the user positions on the paper document 24 within the camera's field of view 70, then it dynamically zooms in the camera until just before it can't "see" it anymore. If the system can't "see" rectangle 66, then it zooms the camera out until it is visible.

Panning is done in a similar way: when one of the sides of the rectangle 66 cannot be seen within the field of view 70, then the camera is panned (by electronic control of the camera's angular orientation) until the side can be seen. Preferably, only certain border regions of the field of view 70, instead processed to determine the presence of the rectangle 66, thereby improving processing speed.

Figure 8 is a flow chart illustrating the processing step in performing the above described panning and zoom operations in accordance with this aspect of the invention.

Initially (step s32), inputs defining the user's adjustment of the (size/location of the) selection rectangle are received, and the accordingly-modified selection rectangle continuously displayed (step s34) as feedback. If it is determined (step s36) that the user has finished selecting (adjusting), and the user chooses to 'copy' what is within the selection rectangle (step s38), the image (IMAGE1) currently within the camera field of view is grabbed (step s40). Then, the displayed (feedback) rectangle is removed (step s42) from the displayed image, and the image (IMAGE2) now within the camera field of view grabbed (step s44). An operation is then performed (step s46) to generate from IMAGE1 and IMAGE2 a difference image.

Using the difference image, the position of the selection rectangle within that image is searched for (step s48), using well known pattern recognition techniques. If the rectangle is not detected (step s50), the camera is zoomed out by a predetermined amount, and step s48 repeated. This continuous in a stepwise fashion until the rectangle is detected at step s50.

Then, an analysis is made (step s54) of whether the rectangle is centred (as between left and right) within the camera field of view, and if not, stepwise panning in the appropriate direction carried out (step s56) until it is detected (step s54) that the rectangle is centred.

Then, an analysis is made (step s58) of whether the rectangle is centred (as between top and bottom) within

the camera field of view, and if not, stepwise panning in the appropriate direction carried out (step s60) until it is detected (step s58) that the rectangle is centred.

Then, an analysis is made (step s62) of whether the rectangle is fully visible within the camera field of view, and if so, stepwise zooming-in carried out (step s64) until it is detected (step s62) that the rectangle is no longer fully visible.

Then, an analysis is made (step s66) of whether the rectangle is fully visible within the camera field of view, and if not, stepwise zooming-out carried out (step s68) until it is detected (step s66) that the rectangle is fully visible. Thus, the point is reached where the rectangle is now just fully visible.

#### Claims

1. An interactive desktop system comprising a work surface, a display device for displaying images in a display area, an image capture device for capturing images present on the work surface, and processor means, coupled to the display device and said capturing means, and adjustment means for adjusting, in response to user action, the field of view of the image capture device.
2. The system according to claim 1, wherein the adjustment means includes means for adjusting, in response to user action, a property of the displayed images.
3. The system according to claim 1 or 2, wherein the adjustment means comprises a frame including a housing in which the image capture device is disposed, the frame being movable whereby the field of view of the image capture device is altered.
4. The system according to claim 3, wherein the adjustment means includes a plurality of sensors for outputting a signal indicative of position; and wherein the processor means includes means for determining the new position, following movement of the housing by a user, of said housing at least in two dimensions parallel to said work surface in dependence upon the signals output by said sensors.
5. The system according to claim 3 or 4, wherein the frame is movable in a direction perpendicular to said work surface, the adjustment means includes a sensor for outputting a signal indicative of vertical position; and wherein the processor means includes means for determining the new position, following movement of the housing by a user, of said housing in three dimensions in dependence upon the signals output by said sensors.
6. The system according to claim 3 or 4, wherein the

frame is rotatable about an axis perpendicular to said work surface, the adjustment means includes a sensor for outputting a signal indicative of angular position; and wherein the processor means includes means for determining the new position, following movement of the housing by a user, of said housing in three dimensions in dependence upon the signals output by said sensors.

7. The system according to claim 4, 5 or 6, wherein the processor means includes means for rendering an altered image for display in dependence upon the determined new position.
8. The system according to any of the preceding claims, wherein said processor means includes means for modifying, in response to a user input, the dimensions of the display area and/or the resolution of the images displayed therein.
9. The system according to any of claims 3 to 8, wherein the processing means includes means, responsive to a user input designating a position on the work surface, for causing movement of said housing whereby the said designated position is moved to within the field of view of said image capture device.
10. The system according to claim 8, wherein the processing means includes means for determining from a captured image whether any of the elements of a predetermined form lie within the field of view of said image capture device, and if none of the elements are present, for causing the image capture device to electronically zoom out.

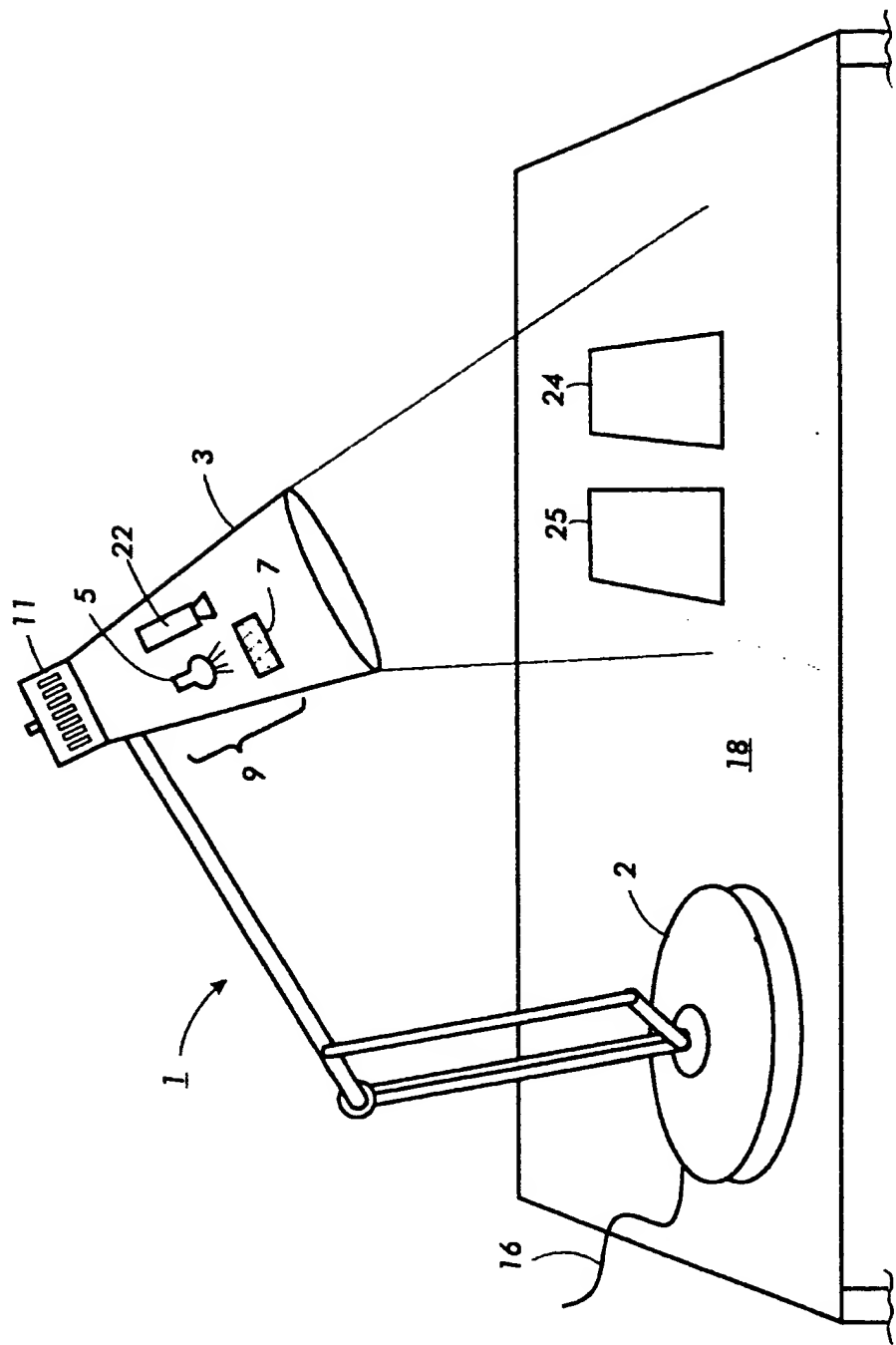


FIG. 1

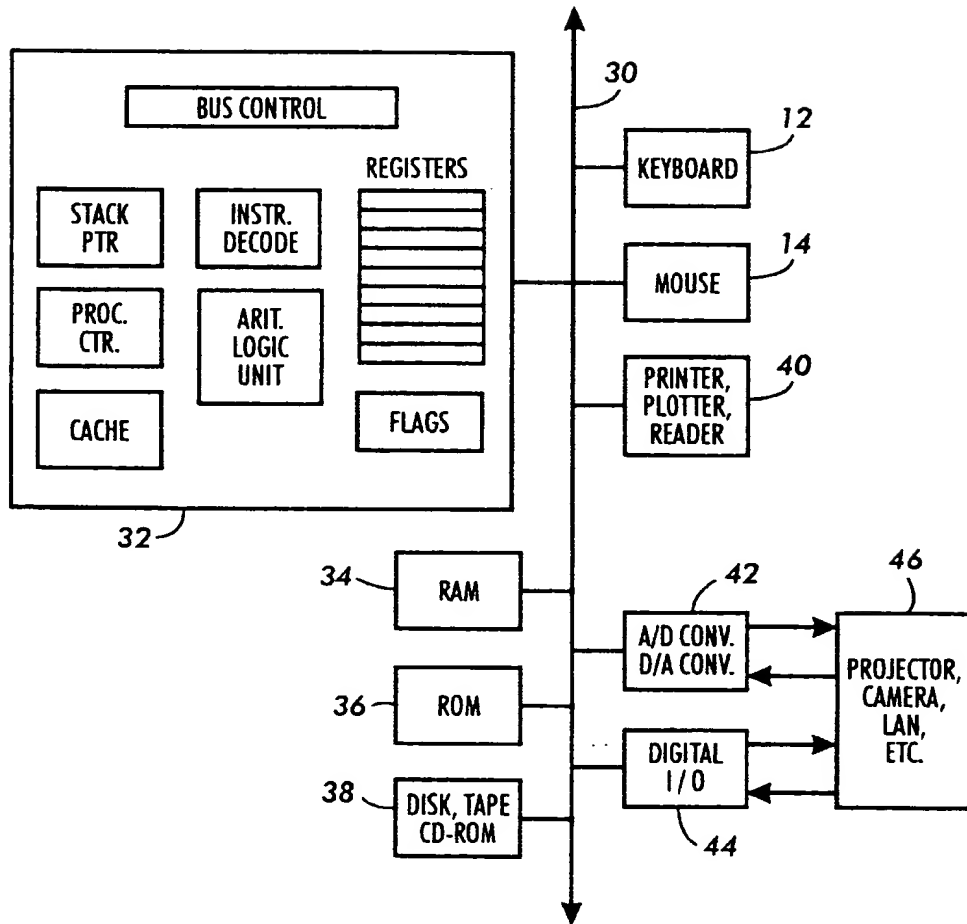


FIG. 2



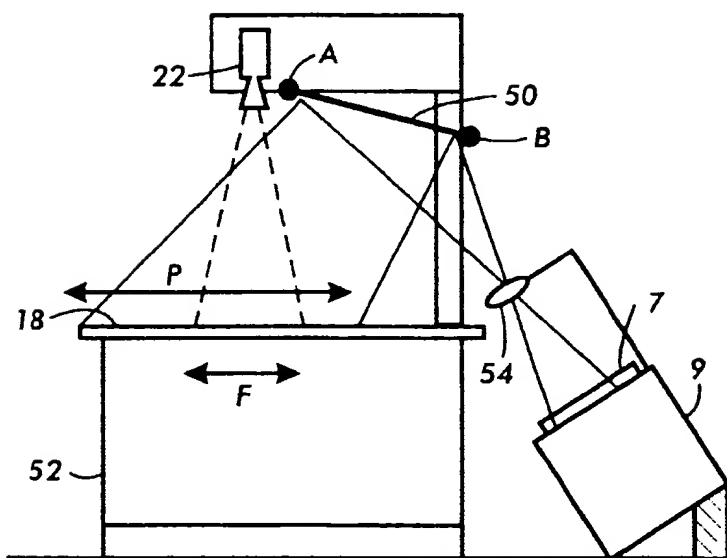


FIG. 3A

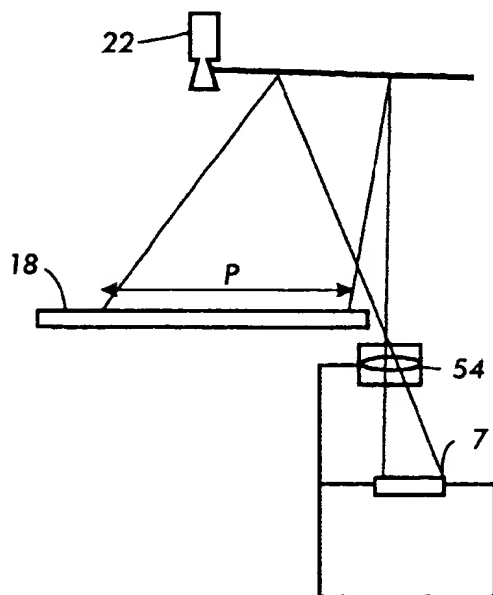
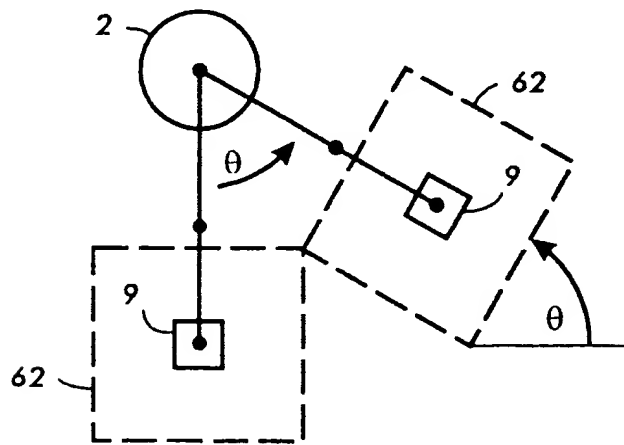
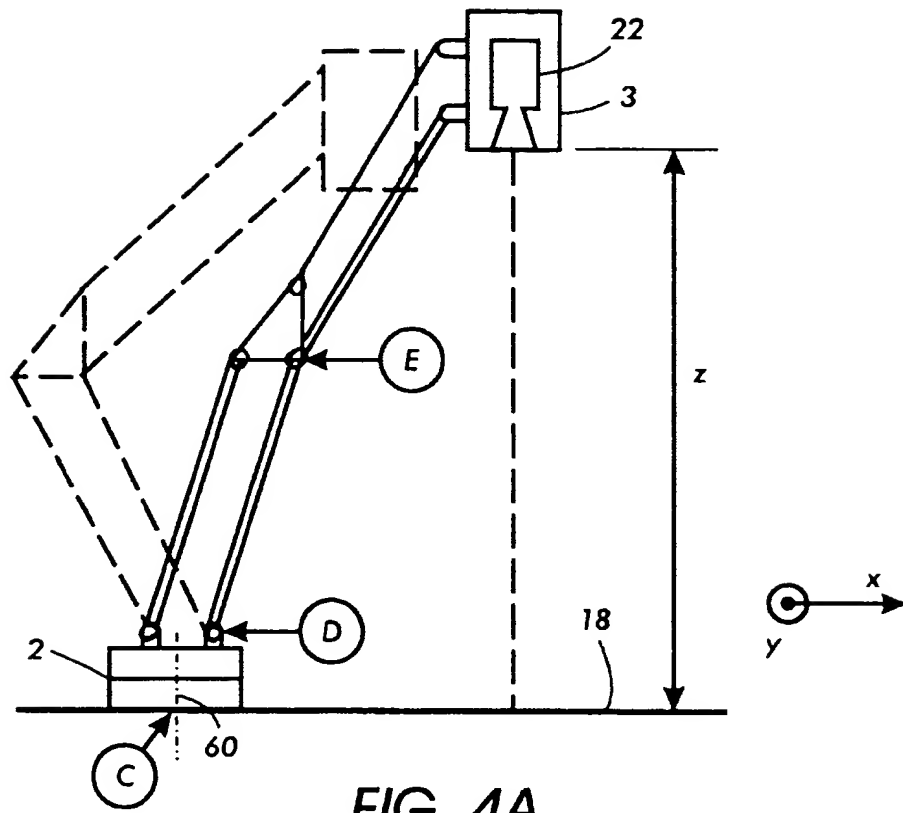


FIG. 3B



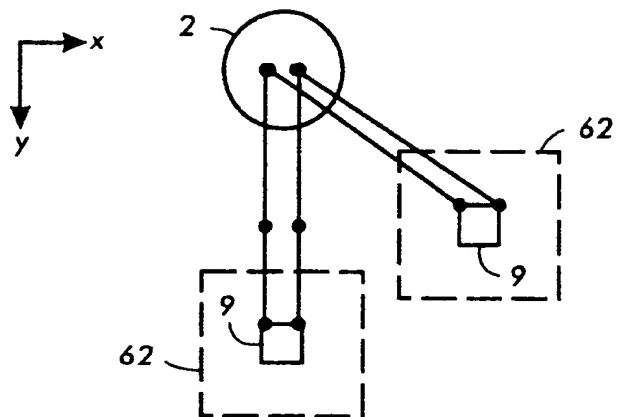


FIG. 5

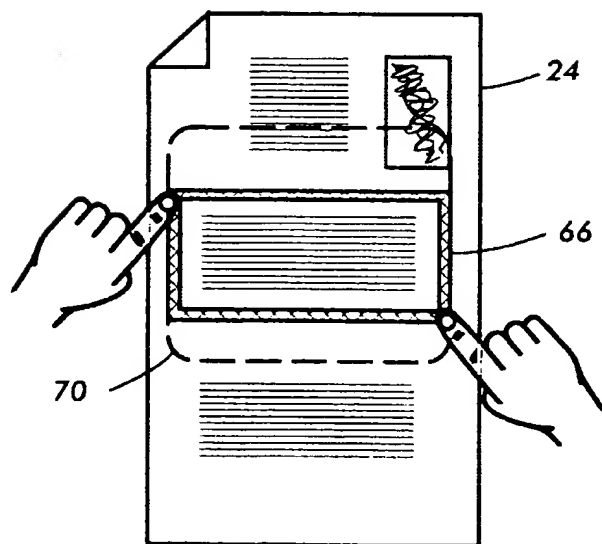


FIG. 7

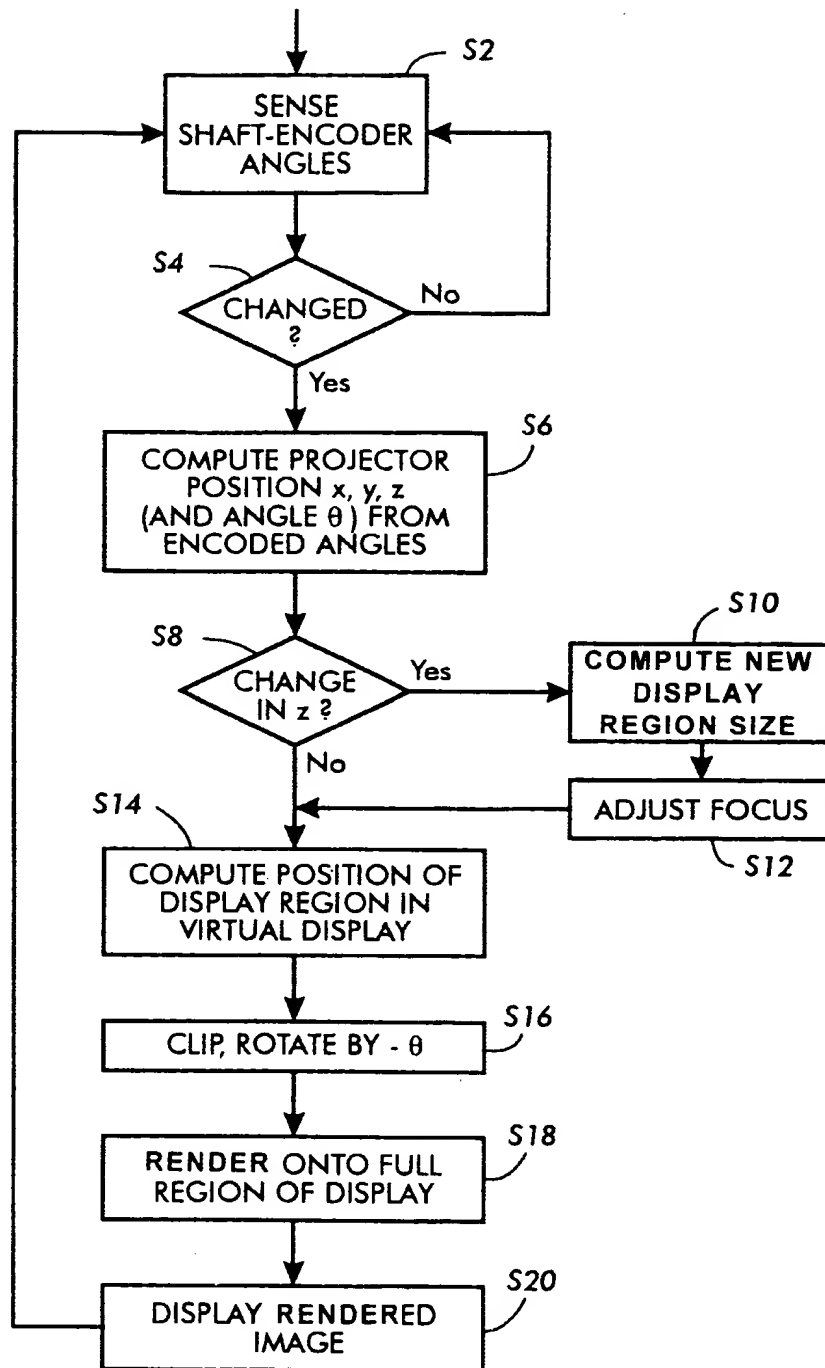


FIG. 6

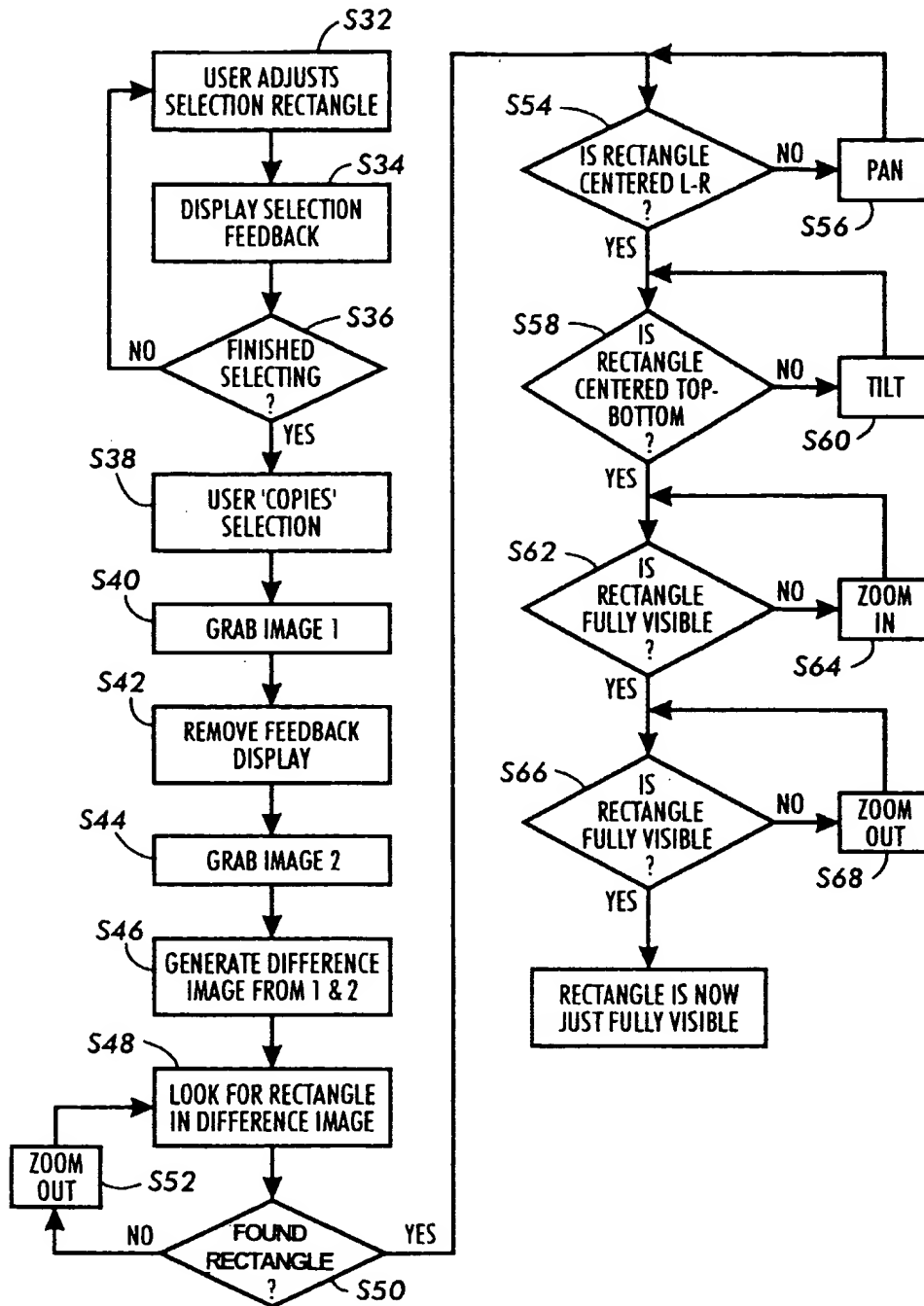


FIG. 8